Plant Biotechnology 32, 299–308 (2015)
DOI: 10.5511/plantbiotechnology.15.1023b

Original Paper

Identification of furostanol glycoside 26-*O*-*β*-glucosidase involved in steroidal saponin biosynthesis from *Dioscorea esculenta*

Masaru Nakayasu[1], Takashi Kawasaki[1], Hyoung Jae Lee[1], Yukihiro Sugimoto[1], Michio Onjo[2], Toshiya Muranaka[3], Masaharu Mizutani[1,*]

[1]Graduate School of Agricultural Science, Kobe University, Kobe, Hyogo 657-8501, Japan; [2]Graduate School of Agriculture, Kagoshima University, Kagoshima, Kagoshima 890-8580, Japan; [3]Department of Biotechnology, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan
*E-mail: mizutani@gold.kobe-u.ac.jp Tel: 81-78-803-5885 Fax: 81-78-803-5884

Received September 25, 2015; accepted October 23, 2015 (Edited by N. Sasaki)

Abstract Steroidal saponins are natural surfactants with various biological activities, and the tubers of *Dioscorea*, known as yam, contain a furostanol glycoside protodioscin and a spirostanol glycoside dioscin, which are valuable saponins required for semi-synthetic production of pharmaceutical steroidal drugs. Steroidal saponins are biosynthesized from cholesterol via several steps of oxygenation and transglycosylation, and a *β*-glucosidase is involved in the hydrolytic conversion from furostanol glycosides to spirostanol glycosides. To investigate steroidal saponin biosynthesis in *Dioscorea* spps, comparative transcriptome analysis of high saponin producers, *D. esculenta* and *D. cayenensis*, and a low producer, *D. alata*, was performed using 454 pyrosequencing. In this study, we isolated and characterized a *β*-glucosidase (DeF26G1) from *D. esculenta*. The *DeF26G1* cDNA encodes a family 1 glucosidase, and the *DeF26G1* transcript was present at high levels in *D. esculenta* but not detected in *D. alata*. The recombinant DeF26G1 protein hydrolyzed the 26-*O*-glycosidic bond of protodioscin to form dioscin, indicating that the *DeF26G1* gene encodes furostanol glycoside 26-*O*-*β*-glucosidase. These results suggested that DeF26G1 is involved in the conversion of furostanol saponins to spirostanol saponins, which seems to be related to biological defense response in the leaves of *Dioscorea* plants.

Key words: Dioscin, *Dioscorea*, *β*-glucosidase, protodioscin, steroidal saponin.

Steroidal saponins consist of the steroidal backbone as an aglycone and the oligosaccharide moiety attached at the C3 and/or C26-hydroxy group. They are natural surfactants with various biological activities such as hemolytic, cytotoxic, anti-inflammatory, antifungal and antibacterial properties (Sparg et al. 2004). The edible tubers of *Dioscorea* species have been reported to contain high amounts of steroidal saponins (sometimes with a yield >2%) as the functional compounds. In particular, the tubers of *Dioscorea* spp. are known to contain furostanol and spirostanol glycosides such as protodioscin and dioscin (Figure 1), respectively, which are valuable saponins used for semi-synthetic production of pharmaceutical steroidal drugs such as anti-inflammatory, androgenic, estrogenic, and contraceptive drugs. Steroidal saponins in *Dioscorea* spp. are biosynthesized from cholesterol via the sequential steps of oxygenation and transglycosylation (Joly et al. 1969a, b; Varma et al. 1969). Cytochrome P450 monooxygenases (CYPs) are likely involved in the oxygenations at the C16, C22, and C26 positions, and UDP-dependent glycosyltransferases (UGTs) also function in the glycosylation at C3 and C26 (Figure 1). But little is known about enzymes and genes for steroidal saponin biosynthesis in *Dioscorea* spp.

In addition, a *β*-glucosidase is involved in the conversion of protodioscin to dioscin. Protodioscin, which contains a glucose unit at the C26-hydroxy group of furostanol, is a precursor of dioscin, and a *β*-glucosidase, furostanol glycoside 26-*O*-*β*-glucosidase (F26G), cleaves the glucose unit to form the E/F spiro ring of dioscin (Figure 1). Purification and cloning of F26G have been reported from several plant species. *Costus speciosus* F26G, which hydrolyzes protogracillin to form gracillin, was purified (Inoue et al. 1996b; Inoue and Ebizuka 1996), and the corresponding *CsF26G* cDNA was isolated (Inoue et al. 1996a). *Avena sativa* avenacosidase, which hydrolyzes avenacosides to 26-degluco-avenacosides, was purified (Grünweller and Kesselmeier 1985; Nisius 1988), and isolation of the

Abbreviations: CYP, Cytochrome P450 monooxygenase; F26G, furostanol glycoside 26-*O*-*β*-glucosidase; UGT, UDP-dependent glycoxyltransferase; GH1, family 1 glycosidase.

This article can be found at http://www.jspcmb.jp/
Published online December 16, 2015

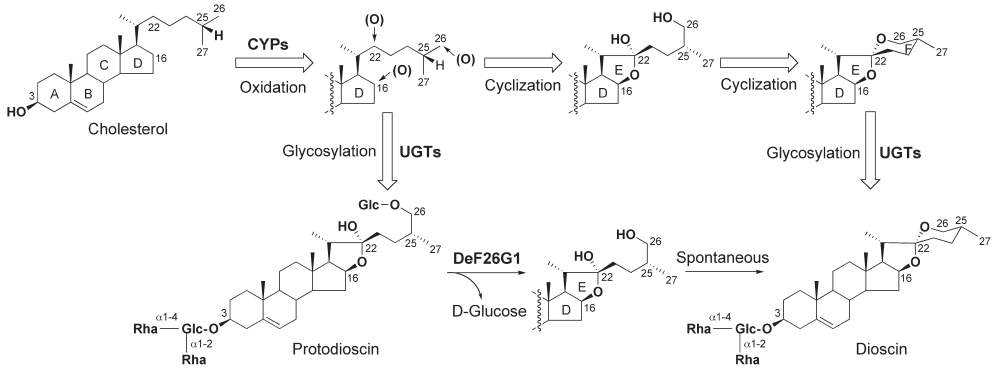


Figure 1. The putative biosynthetic pathway of protodioscin and dioscin in *Dioscorea* spp, and the structures of steroidal saponins, furostanol glycoside protodioscin and spirotanol glycoside dioscin, found in *Dioscorea* spp. Thick arrows indicate unidentified reaction stages. Solid arrows mean reaction steps characterized in this work.

cDNA encoding avenacosidase was reported (Gus-Mayer et al. 1994). These monocotyledonous glucosidases belong to glycosyl hydrolase family 1. In contrast, F26G from *Solanum torvum*, which converts furostanol glycosides torvosides to 26-degluco-torvosides, belongs to glycosyl hydrolase family 3 (Arthan et al. 2006). In the case of *Dioscorea* spp., the F26G activity was reported in *D. pseudojaponica* Yamamoto (Yang et al. 2009), and histochemical analysis of the F26G activity in *D. caucasica* revealed localization of F26G in the thylakoid membrane fraction of the chloroplast (Gurielidze et al. 2004). However, isolation and characterization of the *F26G* gene in *Dioscoreaceae* has so far not been reported.

In this study, we performed comparative transcriptome analysis by using the tubers of high saponin producers, *D. esculenta* and *D. cayenensis*, and a low producer, *D. alata*, as plant materials to investigate steroidal saponin biosynthesis in the yam tubers.

Materials and methods

Plant materials

Dioscorea species used in this study (*D. esculenta*, *D. alata* and *D. cayenensis*) were cultivated in the Experiment Station for Agricultural Science of Kagoshima University. Each part of the cultivated plants was separated, frozen in liquid nitrogen and stored at −80°C.

Chemicals

Protodioscin and dioscin were purchased from Funakoshi Co. (Tokyo, Japan) and *p*-nitrophenyl-*β*-D-glucopyranoside was purchased from NACALAI TESQUE, INC. (Kyoto, Japan).

Extraction of steroidal saponins in plants

Each frozen sample was powdered with a mortar and pestle in the presence of liquid nitrogen. Powdered samples (1 g) were extracted 5 ml of 70% (v/v) aqueous ethanol. After boiled for 1 h at 80°C, the supernatants were then separated following centrifugation at 5000 rpm for 10 min. The supernatant was diluted by 50 times with 70% (v/v) aqueous ethanol and filtered through a 0.22 μm PVDF membrane (Millipore). The filtrate was analyzed by UPLC-ESI-MS.

Crude enzyme in leaves extraction

Frozen sample of leaves was homogenized with a mortar and pestle in 10 ml of ice-cold extraction buffer of crude enzyme (20 mM sodium citrate buffer (pH 6.0), 30 mM 2-mercaptoethanol, 5 mM EDTA, 5% (w/v) PVPP and 0.5 mM PMSF) per g fresh weight material. After centrifugation at 12,000×*g* for 30 min at 4°C, the supernatant was obtained as crude enzyme.

Enzyme assay

The reaction solution (135 μl) of containing 50 mM sodium citrate buffer (pH 7.0) and 100 μM protodioscin was pre-incubated for 5 min at 37°C, and then 15 μl of crude enzyme from leaves or recombinant proteins was added. The reaction was carried out at 37°C for 60 min. The reaction was stopped after by the addition to 150 μl of *n*-BuOH saturated with water including 10 μg genistin per ml as an internal standard. The reaction mixture was vortexed for 3 min and centrifuged at 15,000 rpm for 1 min at room temperature, followed by collection of the upper layer. The above extraction was performed by three times. The obtained *n*-BuOH layer was filtered through a 0.22 μm PTFE membrane (Millipore) and analyzed by UPLC-ESI-MS (Waters). We also measured *β*-glucosidase activity for *p*-nitrophenyl-*β*-D-glucopyranoside. The reaction solution (150 μl) containing 50 mM potassium phosphate buffer (pH 6.0) and 5 mM *p*-nitrophenyl-*β*-D-glucopyranoside was pre-incubated for 5 min at 37°C, and then 50 μl of crude enzyme from leaves of *D. esculenta* or recombinant DeF26G1 was added. The reaction was carried out at 37°C for 15 min. The reaction was stopped after by the addition of 200 μl of 1M Na_2CO_3. The activity was determined spectrophotometrically with liberated *p*-nitrophenol at 405 nm.

RNA extraction

Total RNA of tuber flesh and tuber cortex of *D.esculenta*, *D. alata* and *D. cayenensis* was prepared using the RNAeasy plant mini-kit (QIAGEN, Hilden, Germany), RNase-Free DNase Set

(QIAGEN), and Plant RNA Isolation Reagent (Invitrogen). Each frozen sample was powdered in a mortar with a pestle in the presence of liquid nitrogen. Powdered samples (50 mg) were added to 3 ml of Plant RNA isolation reagent (Invitrogen) and 150 μl of 20% (w/v) PEG. The mixture was incubated at room temperature for 5 min and centrifuged at 15,000 rpm for 10 min at room temperature. The supernatant was added to half volume of 100% EtOH. The mixture was purified by Pink column of RNeasy Plant Mini Kit (QIAGEN) and treated by RNase-Free DNase Set (QIAGEN) according to the instructions from the manufacturer. Total RNA was eluted to microtube with RNase free water. Total RNA of the leaves was extracted from the powdered sample by using TRIzol Reagent (Invitrogen). The obtained RNA was then purified by lithium chloride, followed by sodium acetate.

RT-PCR and TA cloning

Based on *DeF26G1* partial sequence of EST results, the following two primers were designed; DeF26G1_Fw: 5′-GAT CAC TTT GGT TTG TAT GTG GTA-3′, DeF26G1_Rv: 5′-AAT TTA GAG AAC CAT TTT GCA GAC-3′. 100 μg of total RNA was used to synthesize the first strand cDNA (10 ng μl^{-1}) using Transcriptor First Strand cDNA Synthesis Kit (TOYOBO). Reverse transcription was carried out at 37°C, and the PCR was undertaken for 2 min at 95°C, followed by 35 cycles of 20 s at 95°C, 30 s at 49°C, and 60 s at 72°C. PCR products were analyzed on a 1% (w/v) agarose gel, and cDNA fragments of expected size (700-bp) were recovered from the agarose gel using a Qiaquick gel extraction kit (Promega) and cloned into a pMD20 vector (TaKaRa) according to the instructions from the manufacturer.

GS-FLX library construction, sequencing, and functional annotation

Total RNA of the tuber fleshs of *D. esculenta* and *D. alata* was extracted using Plant RNA Isolation Reagent (Invitrogen) and the RNAeasy plant mini-kit (QIAGEN, Hilden, Germany), and DNA contamination was eliminated using RNase-Free DNase Set (QIAGEN). Construction of cDNA libraries for the GS-FLX Titanium System (Roche Diagnostic, Tokyo, Japan), sequencing, cleaning-up and assembly of sequences were performed by Dragon Genomics Center, TakaraBio Inc., Mie Japan). The assembled contigs showing significant similarity to steroidal saponin biosynthetic enzymes such as CYPs, UGTs, and GH1 were identified using the EST Viewer software (Dragon Genomics Center, Takara Bio Co., Ltd., Japan) via the BLASTX search with Arabidopsis CYPs, UGTs, and GH1.

Construction of cDNA library of the tuber fleshs of D. esculenta and isolation of a full-length DeF26G1 cDNA

Total RNA of the tuber fleshs of *D. esculenta* was extracted as described above. cDNA library of the tuber fleshs of *D. esculenta* was constructed using SMART cDNA Library Construction Kit (Clontech Laboratories) according to the manufacture's instruction. Approximately 1,000,000 cDNA-containing phages were screened on nylon filters (Hybond-N, Amersham Biosciences) using an alkaline phosphatase-labeled probe based on *DeF26G1* fragment. A total of 30 positive phages, which are longer than 1000 bp, were isolated and converted to pTriplEx2 plasmids according to the manufacture's instruction (Clontech Laboratories). DNA sequence analysis showed that all three clones were identical. A clone, whose insert was 1,985 bp in length, was selected for further analysis.

The nucleotide sequences were determined using ABI 3130 genetic analyzer (Applied Biosystems) and analyzed using BioEdit, a biological sequence alignment editor (http://www.mbio.ncsu.edu).

Real-Time quantitative RT-PCR analysis

Quantitative RT-PCR was performed with LightCycler®Nano (Roche) using THUNDERBIRDTM SYBR® qPCR Mix (TOYOBO) with the following two sets of primers; *DeF26G1* qPCR Fw: 5′-CAA GCT CTT GAG GAT GAA TAT GGA GGC-3′, *DeF26G1* qPCR Rv: 5′-CTC CAC GGT TCA TTC AAT GTG ATC CAA TAC-3′, *GAPDH* qPCR Fw: 5′- AAT GCT AGC TGC ACC ACC AAC TG-3′, *GAPDH* qPCR Rv: 5′- AAC TGG CAG CTC TTC CAC CTC TC-3′. Cycling was undertaken for 10 min at 95°C, 45 cycles of 10 s at 95°C, 10 s at 60°C, and 15 s at 72°C for amplification, followed by holding for 30 s at 95°C and ramping up from 60 to 95°C at 0.1°C s^{-1} for melting curve analysis. Three biological repeats were analyzed in duplicate. *DeF26G1* gene expression levels were normalized against the values obtained for the *GAPDH* gene, which was used as an internal reference. Data acquisition and analysis were performed using LightCycler®Nano software (Roche).

Expression of recombinant DeF26G1 in E. coli

DeF26G1 cDNA fragments were amplified by PCR using the following three primers containing restriction sites (underlined); *Eco*RI-full-DeF26G1 Fw: 5′-GAA TTC ATG GCC TCA ATA GTC TCT CA- 3′, *Eco*RI-d80-DeF26G1 Fw: 5′-GAA TTC AAG GCC ACT GAA GCA TTT GT- 3′, *Xho*I-DeF26G1 Rv: 5′-CTC GAG CTA GTT TTG AGG CTT TGG- 3′. The PCR was performed for 2 min at 95°C, followed by 30 cycles 30 s at 95°C, 30 s at 59°C, and 2 min at 72°C. The amplified DNA fragments were ligated into the pMD19 vector (TaKaRa) and digested with *Eco*RI and *Xho*I. The DNA fragments were ligated into *Eco*RI–*Xho*I sites of the pGEX4T-1. *E. coli* strain BL21 (DE3) transformed with constructed plasmid was grown at 37°C in LB medium with 50 μg ml^{-1} ampicillin until its OD_{600} reached appropriate 0.5. The recombinant protein expression was induced by adding IPTG to 0.1 mM and continued for 20 h at 18°C. The culture was then centrifuged at 3500 rpm for 30 min at 4°C. The cell pellets were resuspended in 5 ml of cold sonication buffer containing 50 mM sodium phosphate (pH 7.4), 300 mM NaCl and 20% (v/v) glycerol, sonicated using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer (Sigma) typeMS73 at a sound intensity of 200 W cm^{-2}, three times for 30 s each on ice, and centrifuged at 15,000 rpm for 10 min at

4°C. The GST-tagged proteins present in the supernatant were purified using GST Spin Trap columns (GE Healthcare) according to the manufacturer's instructions. After two column washes, the adsorbed proteins were eluted twice in 200 μl of a solution containing 50 mM Tris-HCl (pH 8.0) and 20 mM reduced glutathione, and mixed. The concentration of the purified proteins was determined by Bradford system. The purified recombinant proteins were visualized by SDS-PAGE. The proteins were revealed by staining the gel using Coomassie brilliant blue R-250. The proteins were used for further analyses.

LCMS Analysis of steroidal saponins

LCMS analyses were performed using a system consisting of an Acquity Ultra Performance Liquid Chromatograph (UPLC) (Waters, Milford, MA) and an Acquity quadruple tandem mass spectrometer (TQ Detector) (Waters), and data acquisition and analysis were performed using MassLynx 4.1 software (Waters). Chromatographic column was a Waters ACQUITYTM UPLC HSS T3 column (100×2.1 mm, 1.8 μm). The column temperature was set at 30°C. For each sample, 5 μl was injected. The flow rate was set at 0.2 ml min^{-1}. The mobile phases were water with 0.1% (v/v) formic acid (A) and acetonitrile (B), using a gradient elution of 10–55% B at 0–30 min, 55–75% B at 30–35 min (0–30 min and 30–35 min, linear gradient) for analysis of extracts from plants. While, the mobile phases were 50% (v/v) MeOH in H_2O (A) and 100% MeOH (B), using a gradient elution of 0% B at 0–2 min, 0–100% B at 2–12 min, 100% B at 12–16 min (2–12 min, linear gradient) for analysis of enzymatic activities. The mass spectra of steroidal saponins extracted from plants were obtained in positive ESI mode, while the mass spectra of enzymatic reaction products were detected in negative ESI mode. In ESI conditions, the capillary voltage at 3 kV and sample cone voltage at 80 V were applied. MS scan mode with a mass range of *m*/*z* 400–1100 was used for enzymatic activities analysis using crude enzymes from leaves and recombinant proteins, while SIM mode with *m*/*z* 867 corresponding to representative fragment ion $[M-H]^-$ of dioscin was applied for enzymatic kinetics analysis. On the other hand, SIM mode with *m*/*z* 1031 and *m*/*z* 869 corresponding to representative fragment ion $[M+H-H_2O]^+$ of protodioscin and $[M+H]^+$ of dioscin, respectively, was applied for quantification of steroidal saponins in extracts from plants. The source and desolvation gas temperature were set at 120°C and 350°C, respectively. The nebulizer and desolvation N_2 gas flows were 50 and 550 l/h, respectively.

Biochemical analysis of recombinant d80-DeF26G1

The pH optimum of d80-DeF26G1 was determined using glycine-HCl buffer, sodium citrate buffer and potassium phosphate buffer for the pH ranges 2.0–3.0, 3.0–6.0 and 6.0–8.0, respectively. The activity for each pH was measured as described above. We determined the kinetics parameters of recombinant d80-DeF26G1 in triplicated assays. The activity was assayed using protodioscin at concentration ranging from 10 to 400 μM. Reaction The reaction was carried out at 37°C for 10 min. Extraction and LCMS analysis of the reaction product were performed as described above. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998).

Results

Quantification of Steroidal Saponins in D. esculenta, D. cayenensis and D. alata

The tubers of *Dioscorea* species are known to contain high amounts of steroidal saponins, and we chose three species, *D. esculenta, D. cayenensis*, and *D. alata* as plant materials to study steroidal saponin biosynthesis. *D. esculenta* (known as a lesser yam) is widely distributed and cultivated for foods in Okinawa island, parts of southern Asia, and the Pacific, and the tubers are known as a bitter yam probably due to high contents of saponins. The tubers of *D. esculenta* have also been used traditionally as a medicine in the treatment of various diseases. *D. cayenensis* (known as a yellow Guinea yam) is the most popular and economically important yam in West Africa, and *D. cayenensis* has been reported to contain high amounts of steroidal saponins (Sautour et al. 2004a, b, 2007). *D. alata* (known as a water yam) is widely cultivated as an important tuber crop in tropical and subtropical regions.

First, we analyzed the contents of steroidal saponins in the leaves and tubers of *D. esculenta*, *D. cayenensis*, and *D. alata* by UPLC-ESI-MS (Figure 2). The tubers of *D. esculenta* and *D. cayenensis* contained a spirostanol glycoside dioscin and a furostanol glycoside protodioscin, while their leaves contained only protodioscin but no dioscin was found. In contrast, dioscin and protodioscin were not detected in the leaves and tubers of *D. alata*. Thus, we performed comparative transcriptome analysis of the three *Dioscorea* species to identify the genes responsible for steroidal saponin biosynthesis.

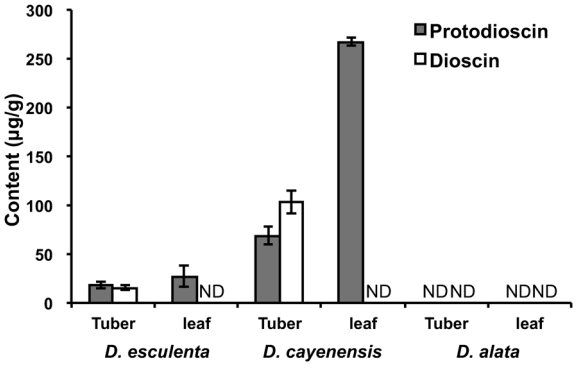


Figure 2. Endogenous contents of protodioscin and dioscin in the tubers and the leaves of *D. esculenta, D. cayenensis, and D. alata*. Bars indicate standard deviation from the mean (n=3). ND indicates not detected.

EST Analysis of Dioscorea species

The comparative transcriptome among the three species was performed by RNA-seq analysis. Total RNA was isolated from the flesh and cortex of the yam tubers, and a transcriptome dataset for each of the extracted RNAs was prepared via 454 pyrosequencing. After cleanup of the sequences and de novo assembly of the cDNA leads, 67,969 contigs were obtained (Figure 3). To identify the candidate genes involved in steroidal saponin biosynthesis, we surveyed three enzyme superfamilies, CYPs, UGTs, and family 1 glycosidases (GH1s) in the transcriptome datasets (Table 1). Among the three superfamilies, GH1 showed a clear difference between high- and low-saponin producers. The lead numbers of GH1 were much higher in the tuber fleshs of *D. esculenta* (2,279 leads/total 578,952 leads) and *D. cayenensis* (1123 leads/121,989 leads) than that in *D. alata* (71 leads/137649 leads). Furthermore, one of the GH1 contigs, which was designated as DeF26G, was highly detected in the tuber fleshs of *D. esculenta* (1526 leads) and *D. cayenensis* (989 leads), indicating that the DeF26G contig account for the vast majority of GH1s expressed in the two species. In contrast, the DeF26G contig was not detected in *D. alata* at all. The DeF26G contig showed the highest similarity in the amino acid sequence to furostanol 26-*O*-*β*-glucosidase of *Costus speciosus* (CsF26G), which hydrolyzes protogracillin to form gracillin (Inoue et al. 1996a). These results, together with the difference in the saponin contents among the three *Dioscorea* species, suggested that DeF26G is involved in steroidal saponin biosynthesis. Therefore, we focused on the characterization of the DeF26G contig in this study.

β-Glucosidase activity in crude enzymes from D. esculenta and D. alata

RNA-seq analysis suggested the existence of the transcript encoding a furostanol 26-*O*-*β*-glucosidase homolog in *Dioscorea*. To examine the possibility, crude enzymes prepared from the leaves of *D. esculenta* and *D. alata* were incubated with protodioscin, and the reaction products were analyzed by UPLC-ESI-MS (Figure 4).

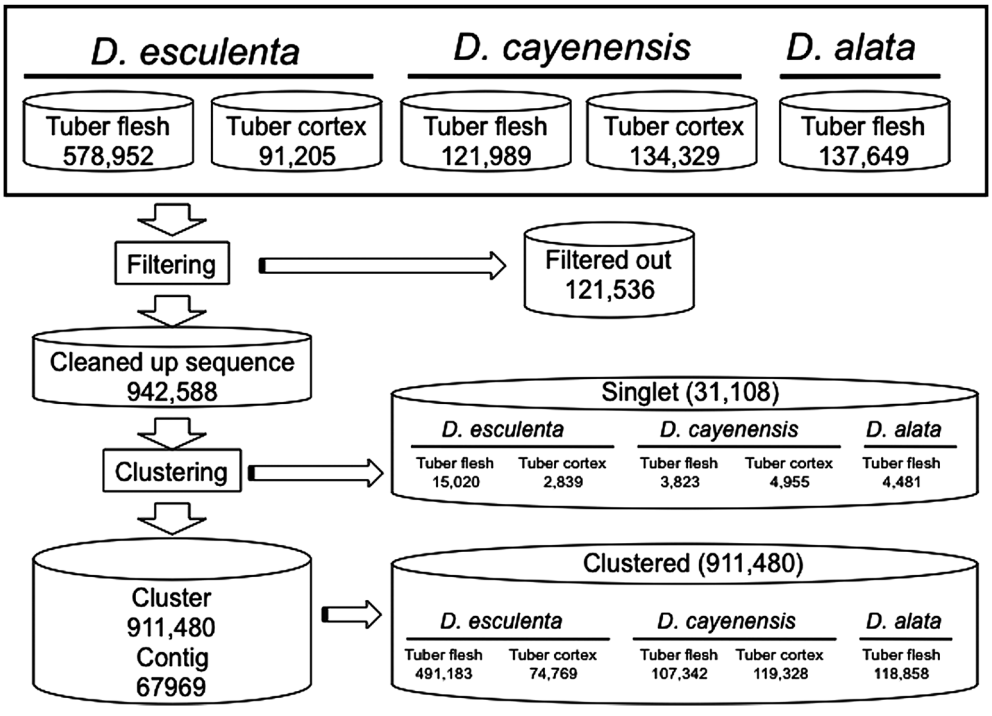


Figure 3. Summary of EST libraries construction of *Dioscorea* spp., *D. esculenta*, *D. cayenensis* and *D. alata*, by RNA-seq analysis.

Table 1. Number of EST reads of putative three enzyme superfamilies, cytochrome P450 monooxygenases (P450s), UDP-dependent glycoxyltransferases (UGTs), and family 1 glycosidases (GH1s), involved in steroidal saponin biosynthesis in *Dioscorea* spp.

Gene Family (contig)	*D. esculenta*		*D. cayenensis*		*D. alata*
	Tuber flesh	Tuber cortex	Tuber flesh	Tuber cortex	Tuber flesh
CYP	4007	722	337	786	299
UGT	1207	243	208	406	232
GH1 (DeF26G)	1526	312	989	27	0
(other GH1)	753	205	134	159	71

After 1 h incubation with protodioscin, the crude enzymes from *D. esculenta* gave a product peak with a retention time at 12.23 min. The product is identical to the authentic compound of dioscin in terms of the retention time and the mass spectra. In contrast, the crude enzymes of *D. alata* did not give the corresponding peak. These results suggested that a *β*-glucosidase hydrolyzing protodioscin to dioscin is present in *D. esculenta* but not in *D. alata*. Thus, the *DeF26G* gene, which is specifically expressed in *D. esculenta*, is likely involved in the hydrolysis of protodioscin.

Isolation of Full-length DeF26G1 cDNA

Based on RNA-seq analysis, a partial cDNA fragment (702-bp) of the *DeF26G* contig was obtained by PCR, and a *DeF26G* cDNA was isolated from a cDNA library of the *D. esculenta* tubers by using the PCR fragment as a probe. The isolated full-length *DeF26G1* cDNA consists of a 1701-bp open-reading frame and a 270-bp 3′-noncoding region. The deduced DeF26G1 protein is a 566 amino-acid precursor polypeptide consisting of a mature protein of 486 amino acid residues and a putative chloroplast transit peptide of 80 amino acid residues at the N-terminus.

The predicted mature polypeptide of DeF26G1 showed the high sequence identity to that of CsF26G (65.5%) and shared 46.5% sequence identity to that of avenacosidase. The phylogenetic analysis of DeF26G1 with other plant GH1s revealed that DeF26G1 was located in the same clade of monocot GH1s including CsF26G and avenacosidase (Supplemental Figure 1). The deduced DeF26G1 contained several sequence motifs that are highly conserved among GH1s (Figure 5). The NEP sequence motif, of which the Glu residue is an acid/base catalyst, was found at residues 257–259, and the sequence ITENG, of which the Glu residue is a catalytic nucleophile of *β*-glucosidases, was also found at residues 468–472 (Jenkins et al. 1995; Keresztessy et al. 1994). The residues involved in the binding of the glycone (glucose)

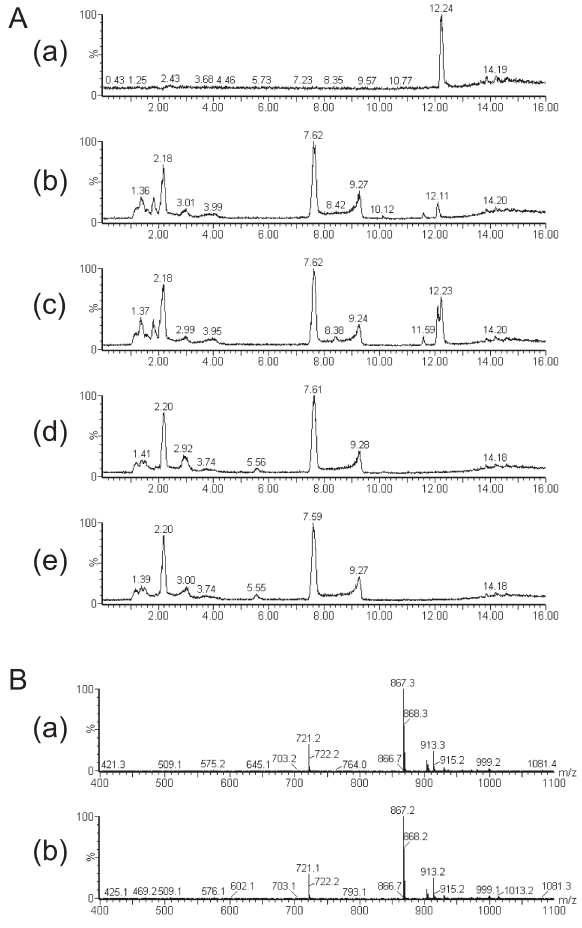


Figure 4. UPLC-ESI-MS analysis of the reaction products from crude enzymes prepared from the leaves of *D. esculenta* and *D. alata* with protodioscin as substrate. (A) Total ion chromatogram [TIC] of the authentic compound of dioscin and the reaction products from crude enzymes. (a) authentic standard of dioscin; (b) control (0 min) of *D. esculenta*; (c) reaction products (1 h) of *D. esculenta*; (d) control (0 min) of *D. alata*; (e) reaction products (1 h) of *D. alata*. (B) (a) Mass spectra of the peak with a retention time at 12.24 min from the authentic compound of dioscin; (b) mass spectra of the product peak with a retention time at 12.23 min after 1 h incubation from the crude enzymes from *D. esculenta*.

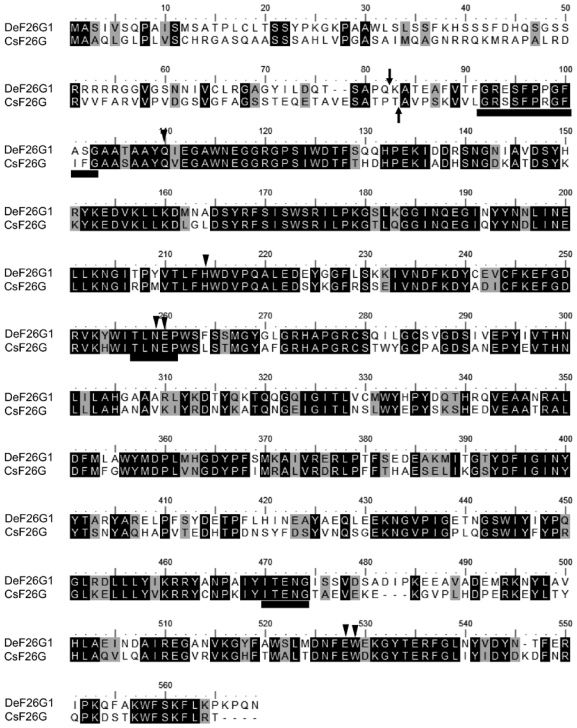


Figure 5. Sequence alignment of amino acid of DeF26G1 and CsF26G. Multiple sequence alignment was performed using the ClustralW multiple alignment analysis tool of BioEdit. Identical and similar amino acid residues are shaded in black and gray respectively. Conserved sequence motifs among family 1 glucosidase (GH1) are underlined. Six amino acids for glucoside recognition are indicated with filled triangles on the sequence. The predicted transit peptide cleavage site of CsF26G and DeF26G1 are indicated with solid arrows. DeF26G1 arrow is the site predicted by the PSORT program (psort.nibb.ac.jp).

moiety are highly conserved in all GH1s (Barrett et al. 1995; Rye and Withers 2000; Saino et al. 2014; Sue et al. 2006; Zechel and Withers 2000), and these residues were also found at Gln-108, His-212, Asn-257, Glu-258, Glu-526, and Trp-527 in the DeF26G1 sequence (Figure 5).

Biochemical characterization of the recombinant DeF26G1

To investigate the potential catalytic activity of DeF26G1, the full-length DeF26G1 protein (full-DeF26G1) and the predicted mature form of DeF26G1 (d80-DeF26G1), of which the 80 amino acid residues at the N-terminus was deleted, were expressed in *Escherichia coli* as a glutathione *S*-transferase fusion protein (Figure 6). The β-glucosidase activity was assayed with protodioscin as a substrate, and the reaction products were analyzed by UPLC-ESI-MS. The reaction product with the recombinant full-DeF26G1 was detected at Rt 12.2 min (Figure 7) and gave an $[M–H]^-$ ion at *m*/*z* 867 (Figure 7). The product is identical to the authentic compound of dioscin in terms of the retention time and the mass spectra. Protodioscin contains the other β-glycosidic bonds of the oligosaccharide moiety attached at the C3-hydroxy position, but these bonds were not hydrolyzed by DeF26G1 because the peak giving an $[M–H]^-$ ion at *m*/*z* 413 corresponding to the aglycone diosgenin was not detected. The recombinant d80-DeF26G1 essentially gave the same results and showed the better hydrolytic activity than full-DeF26G1. Furthermore, a new product peak at Rt 8.42 min was detected with d80-DeF26G1 (Figure 7A (d)). This unknown product gave an $[M–H]^-$ ion at *m*/*z* 885 (Figure 7B (d)), which corresponds to the molecular mass of the 26-degluco form of protodioscin. The results strongly suggested that the product detected at Rt 8.42 min is a hydrolytic product before forming a F-spiro

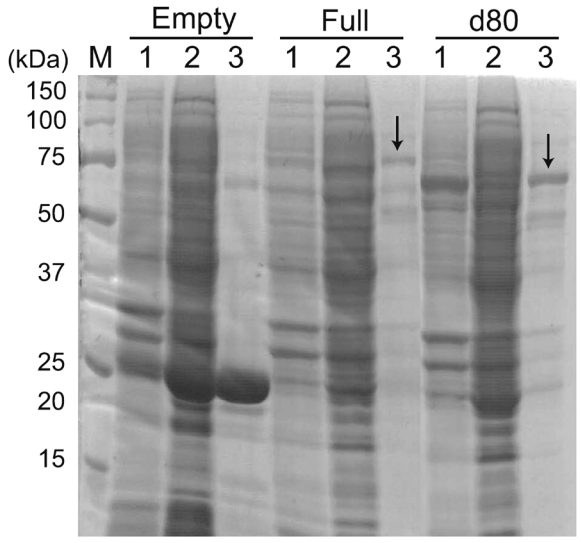


Figure 6. Expression of the recombinant DeF26G1 protein in *E. coli*. *E. coli* BLR cell extracts were analyzed by 10%-SDS-PAGE, and the proteins were visualized by Coomassie Brilliant Blue R-250 staining. Lane M, molecular size markers; lane 1, crude precipitation from the cells containing expression vector; lane 2, crude soluble proteins from the cells; lane 3, the purified GST fusion protein from the crude soluble proteins; Empty; pGEX4T-1 (empty vector), Full and d80; pGEX4T-1 containing Full- and d80-DeF26G1, respectively. Full- and d80-DeF26G1 proteins are indicated with solid arrows.

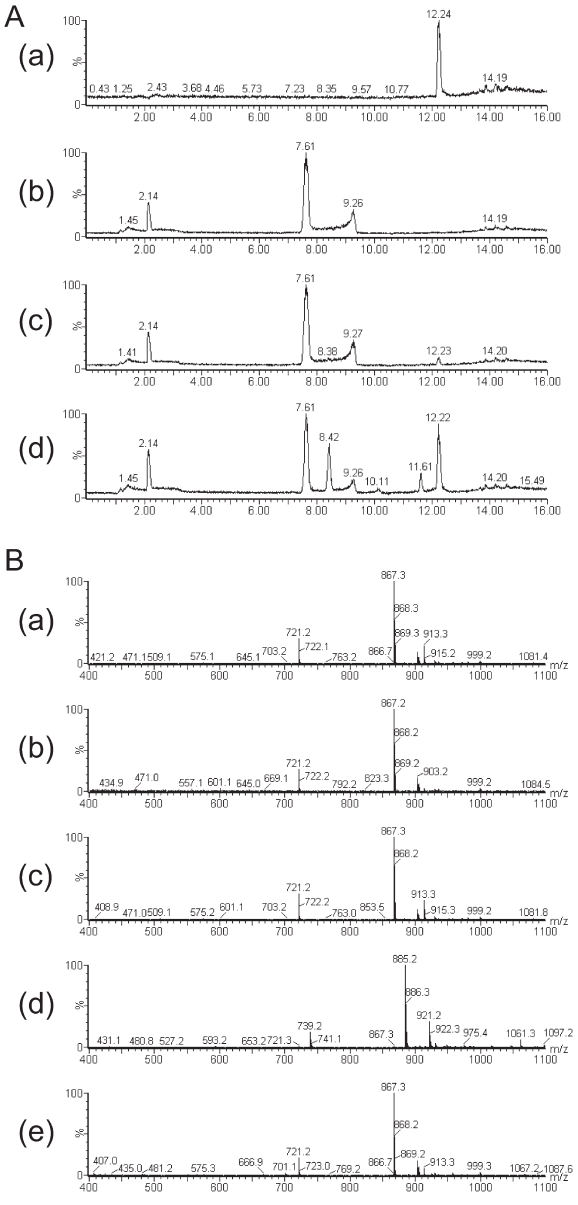


Figure 7. UPLC-ESI-MS analysis of the reaction products from the recombinant GST fusion proteins with protodioscin as substrate. (A) Total ion chromatogram [TIC] of the authentic compound of dioscin and the reaction products from the recombinants. (a) authentic standard of dioscin; (b) the reaction products with the empty vector; (c) the reaction products with the full-length DeF26G1; (d) the reaction products with d80-DeF26G1. (B) (a) Mass spectra of the peak with a retention time at 12.24 min from the authentic compound of dioscin; (b) mass spectra of the product peak with a retention time at 12.2 min from the full-length DeF26G1; (c) mass spectra of the product peak with a retention time at 12.2 min from d80-DeF26G1; (d) mass spectra of the product peak with a retention time at 8.4 min from d80-DeF26G1; (e) mass spectra of the product peak with a retention time at 11.6 min from d80-DeF26G1.

ring and that the 26-degluco form of protodioscin is spontaneously converted to the spirostanol saponin dioscin. Thus, DeF26G1 was able to hydrolyze the 26-*O*-glycosidic bond of protodioscin to form dioscin, clearly indicating that DeF26G1 is a furostanol glycoside 26-*O*-*β*-glucosidase.

The *β*-glucosidase activity of the d80-DeF26G1 toward protodioscin was examined within the range of pH 2–8, and the optimal pH was determined to be pH 7. The optimal pH for DeF26G1 is higher that that of CsF26G (pH 5–5.5). The apparent K_m value for protodioscin was determined to be 140±0.0760 *μ*M (Table 2 and Supplemental Figure 2), and this value is slightly higher than that of CsF26G (50 *μ*M for protogracillin). The d80-DeF26G1 protein did not hydrolyze *p*-nitrophenyl-*β*-D-glucopyranoside, which is a standard artificial substrate for various *β*-glucosidases.

Expression of the DeF26G1 gene in D. esculenta and D. alata

Quantitive real-time-PCR analysis was performed toward total RNA extracted from the tubers and the leaves of *D. esculenta* and *D. alata* (Figure 8). The *DeF26G1* transcript in *D. esculenta* was detected in the tubers and also 6-fold higher in leaves. In contrast, the *DeF26G1* transcript in *D. alata* was not detected at all. These results are consistent with the results of the RNA-seq analysis and the F26G activity with the leaf crude extracts.

Table 2. The biochemical properties of d80-DeF26G1

Parameter	Value
Optimal pH	7.0
K_m [*μ*M]	140±0.0760
k_{cat} [s^{-1}]	0.180±0.00623
k_{cat}/K_m [$mM^{-1}s^{-1}$]	1.28±0.0444

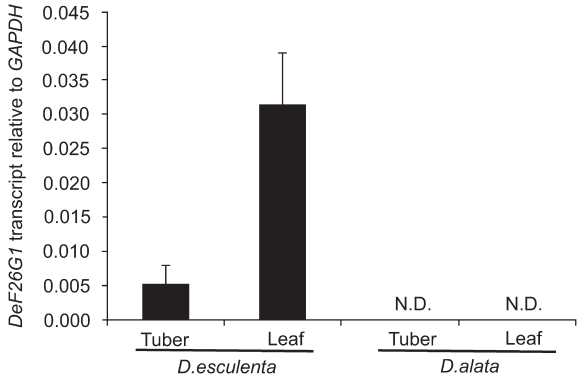


Figure 8. Quantitative RT-PCR analysis of *DeF26G1* gene in different organs of *D. esculenta* and *D. alata*. Values represent the ration between each *DeF26G1* and the corresponding *GAPDH* levels. Bars indicate standard deviation from the mean (n=3). N.D. indicates not detected.

Discussion

We performed comparative transcriptome analysis of three *Dioscorea* species to investigate steroidal saponin biosynthesis in plants. Analysis of endogenous saponin contents in the leaves and tubers revealed a clear contrast between the two *Dioscorea* species, namely, *D. esculenta* as a high saponin producer and *D. alata* as a low saponin producer, and therefore, the transcripts involved in the steroidal saponin biosynthesis were expected to be specifically expressed in *D. esculenta*. The DeF26G contig among the candidate genes (CYPs, UGTs, and GH1) was highly detected in the transcriptome datasets of *D. esculenta* but not of *D. alata* (Table 1), and biochemical characterization clearly indicated that DeF26G1 is a furostanol glycoside 26-*O*-*β*-glucosidase.

In the transcriptome datasets, the DeF26G contig (1526 leads) accounts for 0.31% of total leads (491,183) in the tuber flesh of *D. esculenta* and 0.92% of total leads (107,342) in the tuber fleshs of *D. cayenensis* (Table 1 and Figure 3), and furthermore, the DeF26G contig occupied 67% of total GH1 contigs (2279 leads) in the tuber fleshs of *D. esculenta* and 88% of total GH1 contigs (1123 leads) in the tuber fleshs of *D. cayenensis* (Table 1). These results indicate that DeF26G1 is a predominant *β*-glucosidase in the high saponin producers of *Dioscorea* species. The expression of the *DeF26G1* gene in the leaves of *D. esculenta* was higher than that in the tubers (Figure 8). The endogenous contents of protodioscin, which is a natural substrate of DeF26G1, was higher in the leaves than in the tubers (Figure 2), and thus, the distribution pattern is well consistent with the *DeF26G1* gene expression. It is noteworthy that the leaves of *D. esculenta* and *D. cayenensis* accumulate only a furostanol-type protodioscin but not a spirostanol-type dioscin at all. This suggests that the substrate protodioscin and the corresponding degrading enzyme DeF26G1 are stored in a separate compartment such as subcellular localization or different tissue distribution in the leaves. The full-length DeF26G1 protein contains a putative chloroplast transit peptide of 80 amino acid residues at the N-terminus, suggesting chloroplast localization of DeF26G1. As supporting to these suggestions, Gurielidze et al. (2004) previously reported that the F26G activity is localized in the membrane fraction of the thylakoids and also that furostanol saponins are localized in idioplasts of the epidermis in the leaves of *D. caucasica*. We have shown that DeF26G1 belongs to a GH1, in which various *β*-glucosidases are distributed throughout the plant kingdom. Plant GH1 *β*-glucosidases are involved in defense response against microbials, insects and herbivores, and the pre-toxic glycosides and the degrading GH1s are localized in a separate compartment in plant tissues to prevent poisoning themselves (Ahn et al. 2004, 2006; Morant et al. 2008). Similarly, it is likely

that *Dioscorea* species accumulate furostanol saponins and F26G in a separate compartment of the leaves as a defense mechanism and that, upon cell damage, the hydrolysis of 26-*O*-glycosidic bond of furostanol saponins by F26G may immediately produce a bioactive spirostanol saponins. In contrast to the leaves which accumulate only furostanol saponins, the tubers of *D. esculenta* and *D. cayenensis* contain spirostanol saponin dioscin as well as protodioscin. The different distribution of furostanol and spirostanol saponins in leaves and tubers suggested the presence of unresolved mechanisms by which the organ specific biosynthesis and transport of steroidal saponins are regulated in *Dioscorea* plants. To address the question, we need to identify the biosynthetic genes such as CYPs, UGTs, and transporters and to characterize their tissue specific expression in planta.

In conclusion, we performed comparative transcriptome analysis of *Dioscorea* species, together with chemical analysis of steroidal saponin contents, and identified DeF26G1 as a furostanol glycoside 26-*O*-β-glucosidase, which is involved in the conversion of protodioscin to dioscin (Figure 1). These approaches are powerful to isolate the candidate genes involved in the steroidal saponin biosynthesis. Several candidate genes (CYPs and UGTs) have been found in the transcriptome datasets used in this study (Table 1 and Figure 1), and their biochemical characterization will unravel the molecular mechanism of steroidal saponin biosynthesis in plants.

Acknowledgements

We thank Yasuo Yamauchi (Kobe University) for helpful discussion. This study was supported by the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN), Japan.

References

Ahn YO, Mizutani M, Saino H, Sakata K (2004) Furcatin hydrolase from *Viburnum furcatum* Blume is a novel disaccharide-specific acuminosidase in glycosyl hydrolase family 1. *J Biol Chem* 279: 23405–23414

Ahn YO, Saino H, Mizutani M, Shimizu BI, Sakata K (2007) Vicianin hydrolase is a novel cyanogenic β-glycosidase specific to β-vicianoside (6-*O*-α-L-arabinopyranosyl-β-D-glucopyranoside) in seeds of *Vicia angustifolia*. *Plant Cell Physiol* 48: 938–947

Arthan D, Kittakoop P, Esen A, Svasti J (2006) Furostanol glycoside 26-*O*-β-glucosidase from the leaves of *Solanum torvum*. *Phytochemistry* 67: 27–33

Barrett T, Suresh CG, Tolley SP, Dodson EJ, Hughes MA (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* 3: 951–960

Grünweller S, Kesselmeier J (1985) Characterization of a membrane bound β-glucosidase responsible for the activation of oat leaf saponins. *Phytochemistry* 24: 1941–1943

Gurielidze K, Gogoberidze M, Dadeshidze I, Vardosanidze M, Djaoshvili M, Lomkatsi N (2004) Tissue and subcellular localization of oligofurostanosides and their specific degrading β-glucosidase in *Dioscorea caucasica* Lipsky. *Phytochemistry* 65: 555–559

Gus-Mayer S, Brunner H, Schneider-Poetsch HA, Rüdiger W (1994) Avenacosidase from oat: Purification, sequence analysis and biochemical characterization of a new member of the BGA family of β-glucosidases. *Plant Mol Biol* 26: 909–921

Hernandez A, Ruiz MT (1998) An EXCEL template for caluculation of enzyme kinetics parameters by non-linear regression. *Bioinformatics* 14: 227–228

Inoue K, Ebizuka Y (1996) Purification and characterization of furostanol glycoside 26-*O*-β-glucosidase from *Costus speciosus* rhizomes. *FEBS Lett* 378: 157–160

Inoue K, Shibuya M, Yamamoto K, Ebizuka Y (1996a) Molecular cloning and bacterial expression of a cDNA encoding furostanol glycoside 26-*O*-β-glucosidase of *Costus speciosus*. *FEBS Lett* 389: 273–277

Inoue K, Shimomura K, Kobayashi S, Sankawa U, Ebizuka Y (1996b) Conversion of furostanol glycoside to spirostanol glycoside by β-glucosidase in *Costus speciosus*. *Phytochemistry* 41: 725–727

Jenkins J, Leggio LL, Harris G, Pickersgill R (1995) β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Lett* 362: 281–285

Joly RA, Bonner J, Bennett RD, Heftmann E (1969a) Conversion of an open-chain saponin to dioscin by a *Dioscorea floribunda* homogenate. *Phytochemistry* 8: 1445–1447

Joly RA, Bonner J, Bennett RD, Heftmann E (1969b) The biosynthesis of steroidal sapogenins in *Dioscorea floribunda* from doubly labelled cholesterol. *Phytochemistry* 8: 1709–1711

Keresztessy Z, Kiss L, Hughes MA (1994) Investigation of the active site of the cyanogenic β-D-glucosidase (linamarase) from *Manihot esculenta* Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch Biochem Biophys* 315: 323–330

Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sánchez-Pérez R, Møller BL, Bak S (2008) β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69: 1795–1813

Nisius A (1988) The stromacentre in *Avena* plastids: An aggregation of β-glucosidase responsible for the activation of oat-leaf saponins. *Planta* 173: 474–481

Rye CS, Withers SG (2000) Glycosidase mechanisms. *Curr Opin Chem Biol* 4: 573–580

Saino H, Shimizu T, Hiratake J, Nakatsu T, Kato H, Sakata K, Mizutani M (2014) Crystal structures of β-primeverosidase in complex with disaccharide amidine inhibitors. *J Biol Chem* 289: 16826–16834

Sautour M, Mitaine-Offer AC, Lacaille-Dubois MA (2007) The *Dioscorea* genus: A review of bioactive steroid saponins. *J Nat Med* 61: 91–101

Sautour M, Mitaine-Offer AC, Miyamoto T, Dongmo A, Lacaille-Dubois MA (2004a) A new steroidal saponin from *Dioscorea cayenensis*. *Chem Pharm Bull* 52: 1353–1355

Sautour M, Mitaine-Offer AC, Miyamoto T, Dongmo A, Lacaille-Dubois MA (2004b) Antifungal steroid saponins from *Dioscorea cayenensis*. *Planta Med* 70: 90–92

Sparg S, Light ME, Van Staden J (2004) Biological activities and distribution of plant saponins. *J Ethnopharmacol* 94: 219–243

Sue M, Yamazaki K, Yajima S, Nomura T, Matsukawa T, Iwamura

H, Miyamoto T (2006) Molecular and structural characterization of hexameric *β*-D-glucosidases in wheat and rye. *Plant Physiol* 141: 1237–1247

Varma KR, Wickramasinghe JAF, Caspi E (1969) Biosynthesis of Biosynthesis of plant sterols. IX. The mode of oxygenation at carbon atom 26 in the formation of sapogenins from cholesterol. *J Biol Chem* 244: 3951–3957

Yang DJ, Lu TJ, Hwang LS (2009) Effect of endogenous glycosidase on stability of steroidal saponins in Taiwanese yam (*Dioscorea pseudojaponica* Yamamoto) during drying processes. *Food Chem* 113: 155–159

Zechel DL, Withers SG (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res* 33: 11–18

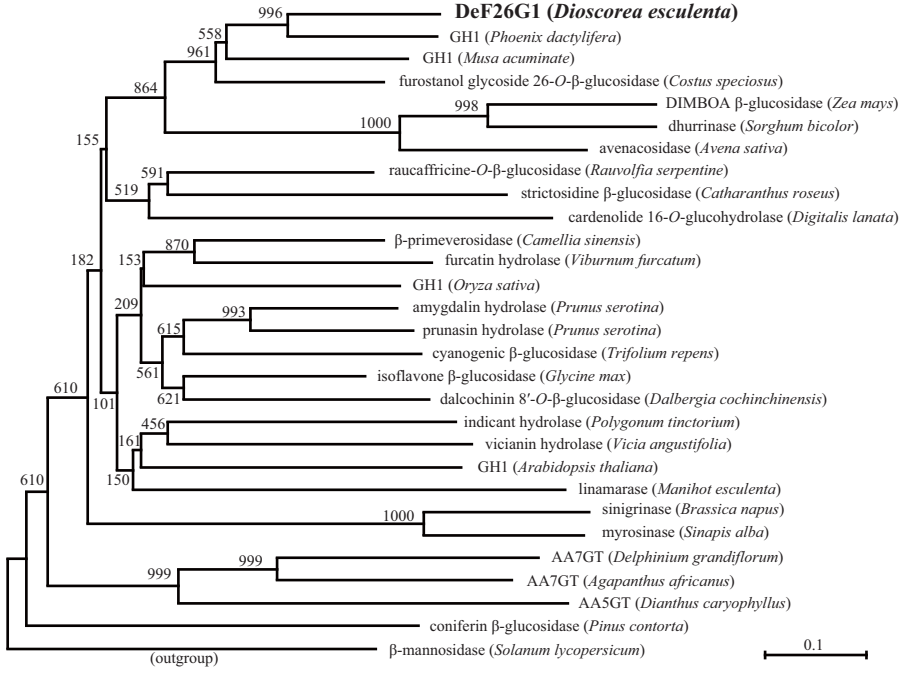
Supplemental Figure 1

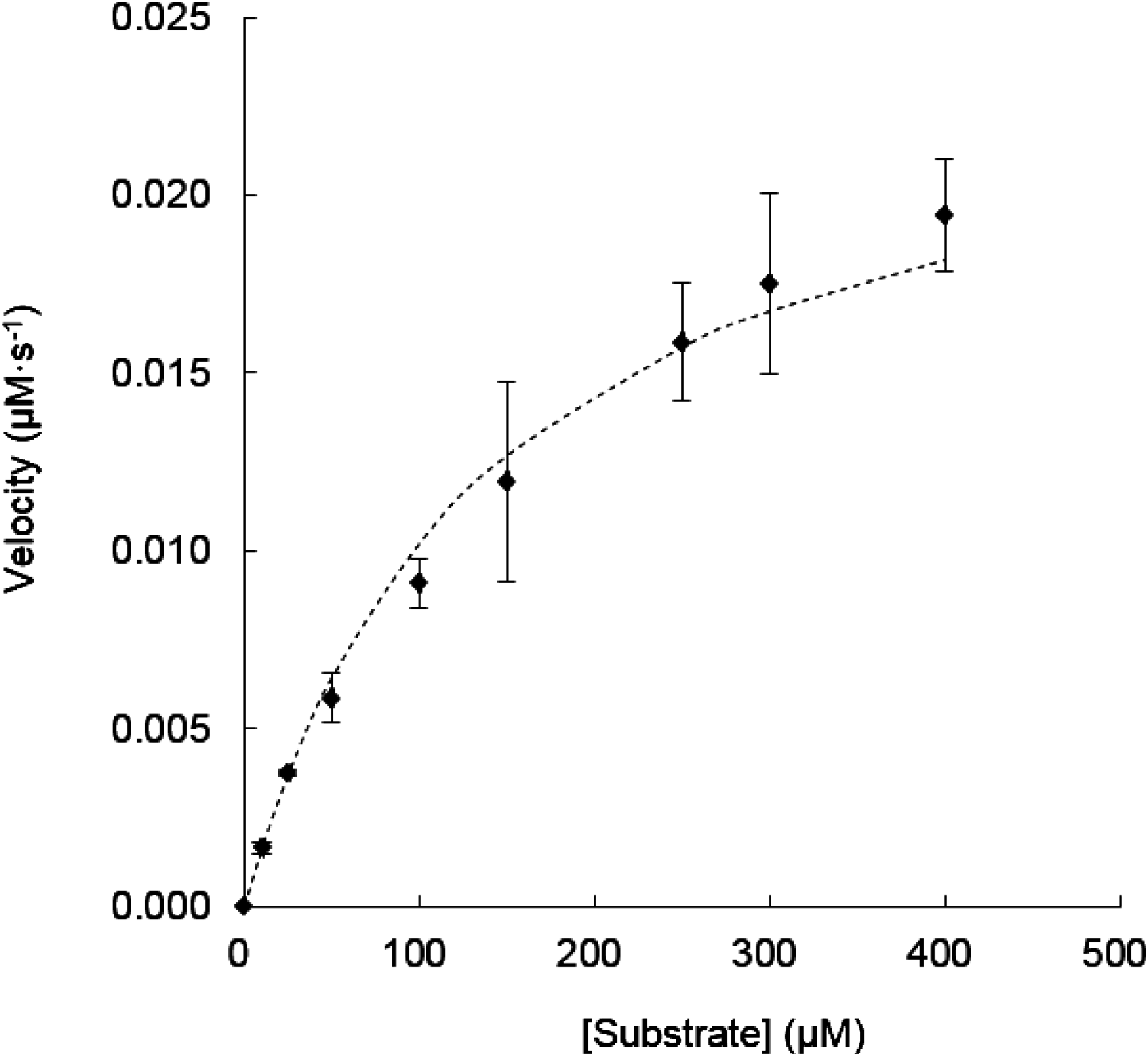
A neighbor-joining phylogenetic tree of glycosyl hydrolase family 1 (GH1) members from various plants. The entire amino acid sequences and their functions of the GH1 members were obtained from the published studies. The Genbank ID numbers of the GH1 members are as follows: GH1 (*Phoenix dactylifera*), XP_008799933; GH1 (*Musa acuminate*), XP_009416645; b-primeverosidase (*Camellia sinensis*), Q7X9A9; furcatin hydrolase (*Viburnum furcatum*), Q75W17; AA5GT (*Dianthus caryophyllus*), BAJ33501; AA7GT (*Delphinium grandiflorum*) BAJ33502; AA7GT (*Agapanthus africanus*), BAM29304; vicianin hydrolase (*Vicia angustifolia*), A2SY66; GH1 (Oryza-sativa), Q7XKV5; amygdalin hydrolase (*Prunus serotina*), Q40984; prunasin hydrolase (*Prunus serotina*), Q9M5X4; indicant hydrolase (*Polygonum tinctorium*), Q9XJ67; GH1 (*Arabidopsis thaliana*), Q8GY78; DIMBOA b-glucosidase (*Zea mays*), P49235; dhurrinase (*Sorghum bicolor*), Q41290; isoflavone b-glucosidase (*Glycine max*), BAF34333; coniferin b-glucosidase (*Pinus contorta*), Q9ZT64, dalcochinin 8′-*O*-b-glucosidase (*Dalbergia cochinchinensis*), Q9SPK3; raucaffricine-*O*- b-glucosidase (*Rauvolfia serpentine*), Q9SPP9; strictosidine b-glucosidase (*Catharanthus roseus*), Q9M7N7; linamarase (Manihot esculenta), Q41172; cyanogenic b-glucosidase *(Trifolium repens)*, P26205; b-mannosidase *(Solanum lycopersicum),* AAL37714; sinigrinase (*Brassica napus*), Q00326; myrosinase (*Sinapis alba*), P29736; cardenolide 16-*O*-glucohydrolase *(Digitalis lanata)*, CAB38854. The statistical significance of the NJ tree topology was evaluated by bootstrap analysis with 1000 iterative tree constructions. The scale indicates the evolutionary distances of the base substitution per site, estimated by the Kimura's two-parameter method.

Supplemental Figure 2

d80-DeF26G1 recombinant enzyme activity curves.

Enzyme activities were measured with substrate concentrations up to 400 μM protodioscin. Michaelis–Menten curve (featuring K_m of 140 μM and V_{max} of 0.0245 $\mu M \cdot s^{-1}$) was fitted to the values obtained. Kinetic parameters were determined by non-linear regression with ANEMONA (Hernandez and Ruiz, 1998).


DeF26G1 (*Dioscorea esculenta*)
GH1 (*Phoenix dactylifera*)
GH1 (*Musa acuminate*)
furostanol glycoside 26-*O*-β-glucosidase (*Costus speciosus*)
DIMBOA β-glucosidase (*Zea mays*)
dhurrinase (*Sorghum bicolor*)
avenacosidase (*Avena sativa*)
raucaffricine-*O*-β-glucosidase (*Rauvolfia serpentine*)
strictosidine β-glucosidase (*Catharanthus roseus*)
cardenolide 16-*O*-glucohydrolase (*Digitalis lanata*)
β-primeverosidase (*Camellia sinensis*)
furcatin hydrolase (*Viburnum furcatum*)
GH1 (*Oryza sativa*)
amygdalin hydrolase (*Prunus serotina*)
prunasin hydrolase (*Prunus serotina*)
cyanogenic β-glucosidase (*Trifolium repens*)
isoflavone β-glucosidase (*Glycine max*)
dalcochinin 8′-*O*-β-glucosidase (*Dalbergia cochinchinensis*)
indicant hydrolase (*Polygonum tinctorium*)
vicianin hydrolase (*Vicia angustifolia*)
GH1 (*Arabidopsis thaliana*)
linamarase (*Manihot esculenta*)
sinigrinase (*Brassica napus*)
myrosinase (*Sinapis alba*)
AA7GT (*Delphinium grandiflorum*)
AA7GT (*Agapanthus africanus*)
AA5GT (*Dianthus caryophyllus*)
coniferin β-glucosidase (*Pinus contorta*)
β-mannosidase (*Solanum lycopersicum*)
(outgroup)
0.1


0.025
0.020
0.015
0.010
0.005
0.000
Velocity (μM·s-1)
0
100
200
300
400
500
[Substrate] (μM)